

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-114616

(43)Date of publication of application : 02.05.1995

(51)Int.CI.

G06K 9/00
G06F 19/00

(21)Application number : 05-262100

(71)Applicant : HITACHI LTD

(22)Date of filing : 20.10.1993

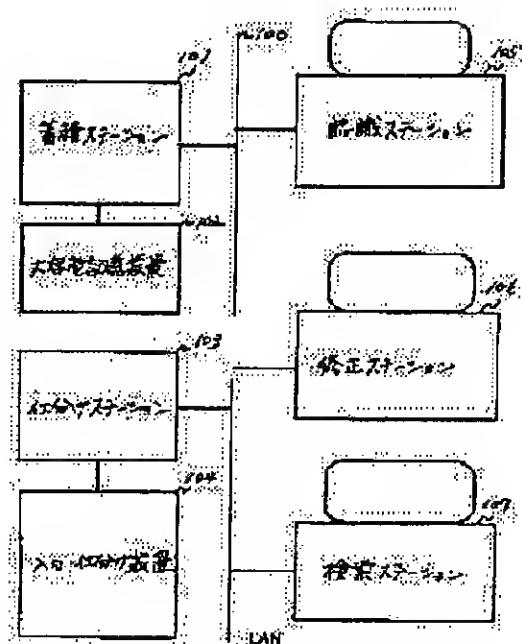
(72)Inventor : SHIMA YOSHIHIRO
HATANO HIDEKAZU
MARUKAWA KATSUMI
KOYA MASASHI
NAKAJIMA KAZUKI
KADOTA AKIZO
KURINO KIYOMICHI
SUGIMOTO TAKEYUKI

(54) SLIP DOCUMENT INFORMATION SYSTEM

(57)Abstract:

PURPOSE: To examine the adequacy of entry contents by discriminating the kinds of the formats of plural slip documents, and automatically setting a read area and performing character recognition.

CONSTITUTION: The slip document information system is constituted by connecting a sorting station 103, a storage station 101, a recognition station 105, a correction station 106, and a retrieval station 107 to a network 100. Consequently, plural slip documents can be read. Further, the adequacy of entry contents is examined and the entry contents of a slip can automatically be inspected. Paper documents can automatically be sorted to reduce the sorting operation.



LEGAL STATUS

[Date of request for examination] 28.03.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C) 1998,2000 Japanese Patent Office

E5813

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平7-114616

(43)公開日 平成7年(1995)5月2日

(51) Int.Cl. [*] G 06 K 9/00 G 06 F 19/00	識別記号 K 8623-5L	府内整理番号 8724-5L	F I G 06 F 15/ 22	技術表示箇所
--	-------------------	-------------------	----------------------	--------

審査請求 未請求 請求項の数11 OL (全 19 頁)

(21) 出願番号 特願平5-262100	(71) 出願人 株式会社日立製作所 東京都千代田区神田駿河台四丁目6番地
(22) 出願日 平成5年(1993)10月20日	(72) 発明者 船 好博 東京都国分寺市東恋ヶ窪1丁目280番地 株式会社日立製作所中央研究所内
	(72) 発明者 羽田野 英一 東京都国分寺市東恋ヶ窪1丁目280番地 株式会社日立製作所中央研究所内
	(72) 発明者 丸川 勝美 東京都国分寺市東恋ヶ窪1丁目280番地 株式会社日立製作所中央研究所内
	(74) 代理人 弁理士 小川 勝男

最終頁に続く

(54)【発明の名称】 伝票文書情報システム

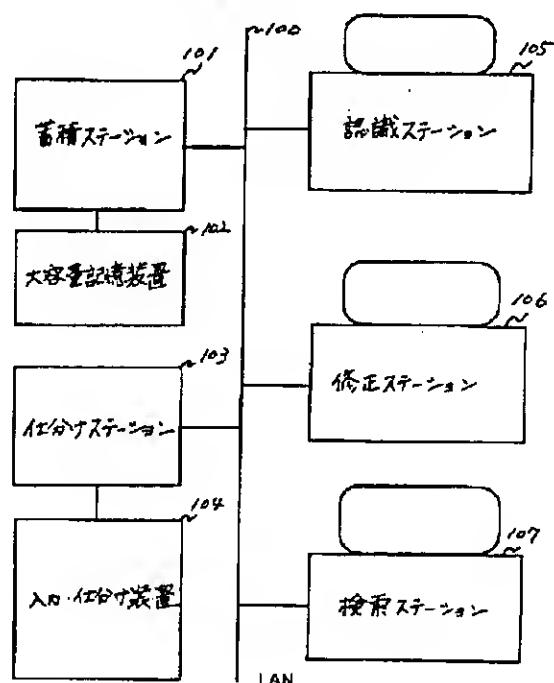
(57)【要約】

【目的】複数の伝票文書の様式の種類を識別し、読み取り領域を自動的に設定し文字認識を行い、記載内容の妥当性を検定することにある。

【構成】伝票文書情報システムの構成は、ネットワーク100に、仕分けステーション103、蓄積ステーション101、認識ステーション105、修正ステーション106、検索ステーション107が接続されている。

【効果】複数の伝票文書を読み取ることができる。また、記載内容の妥当性を検定しており、伝票の記載内容を自動的に点検することができる。紙文書の仕分けを自動的に行なうことができ、仕分け作業を低減するという効果がある。

図 1



【特許請求の範囲】

【請求項1】診療報酬請求明細書の少なくとも画像入力、蓄積、検索、及び内容点検を行なう伝票文書情報システムにおいて、当該明細書の様式の種類を識別する手段と、当該様式の種類に応じて読み取り領域を抽出する手段と、読み取り領域を設定し文字認識を行なう手段と、当該認識結果のデータを基に記載内容の妥当性を検定する手段とを有することを特徴とする伝票文書情報システム。

【請求項2】請求項1記載において、前記明細書の様式の種類を識別する手段は、明細書の種類をキー入力する手段と、様式の種類を予め辞書データとして格納する手段と、入力された明細書文書より様式データを抽出する手段と、当該辞書データと抽出した様式データとを照合する手段を有し、当該キー入力した明細書の種類により、辞書データとの照合を限定することを特徴とする伝票文書情報システム。

【請求項3】紙文書の少なくとも画像入力、蓄積、及び検索を行なう伝票文書情報システムにおいて、伝票文書の種類を識別し文字認識を行なう手段と、認識結果によって整理番号を紙文書に打番する手段と、当該文書の種類に従って紙文書の仕分けを行なう手段とを有することを特徴とする伝票文書情報システム。

【請求項4】紙文書の少なくとも画像入力、蓄積、及び検索を行なう伝票文書情報システムにおいて、伝票文書の種類を識別し文字認識を行なう手段と、当該認識結果に従ってバーコードの印字を紙文書の表面に行なう手段とを有することを特徴とする伝票文書情報システム。

【請求項5】紙文書の少なくとも画像入力、蓄積、表示、及び検索を行なう文書情報システムにおいて、文書画像の複数領域を表示領域として設定する手段と、当該複数領域内において読み取り領域を抽出する手段と、当該表示領域の画像と認識結果を画面に同時に表示する手段と、認識結果を画面を介して修正するときに読み取り領域の画像を拡大表示する手段とを有することを特徴とする文書情報システム。

【請求項6】紙文書の少なくとも画像入力、蓄積、検索、及び内容点検を行なう伝票文書情報システムにおいて、伝票文書の様式の種類を識別する手段と、当該様式の種類に応じて守秘内容の記載された領域を抽出する手段と、当該守秘領域の画像をマスクして表示する手段と、使用者に付与されたパスワードのレベルにより当該マスクの解除、及び設定を行なう手段とを有することを特徴とする伝票文書情報システム。

【請求項7】紙文書の画像入力を行なう伝票文書情報システムにおいて、入力された伝票文書画像から枠線を抽出する手段と、標準的な伝票文書の枠線を登録する手段と、抽出した枠線と類似する枠線を登録データから選択する手段と、当該選択した登録枠線と抽出枠線を画面に同時に表示する手段とを有し、伝票文書を登録すること

を特徴とする伝票文書情報システム。

【請求項8】紙文書の少なくとも画像入力、蓄積、検索、及び内容点検を行なう伝票文書情報システムにおいて、伝票文書の読み取り領域を抽出する手段と、読み取り領域内において文字認識を行なう手段と、認識結果の数値データが予め設定された範囲内にあるかどうかを判定する手段と、当該判定結果に従って伝票文書を蓄積する手段とを有し、当該範囲内にあると判定された伝票文書画像を選択し、当該画像のみを蓄積、及び検索することを特徴とする伝票文書情報システム。

【請求項9】ネットワークを介して紙文書の仕分けを行なう文書情報システムにおいて、紙文書の表面画像を入力する入力ステーションと、入力した画像の種類を識別する認識ステーションと、識別結果に従って紙文書を区分する仕分けステーションと、画像を保管する蓄積ステーションを備え、ネットワークを経由して画像データ、当該種類データ、並びに仕分け制御情報を伝送することを特徴とする文書情報システム。

【請求項10】紙文書の画像を入力し、当該紙文書画像を認識して種類を識別する仕分けステーションであつて、紙文書を設定するホッパと仕分け後の紙文書を格納するスタッカを有し、スタッカが一杯になると警告音を発生し仕分け処理を一時停止することを特徴とする仕分けステーション。

【請求項11】文書の少なくとも画像入力、蓄積、検索、及び内容点検を行なう文書情報システムにおいて、文書画像とその整理番号を格納する手段と、文書の整理番号を指定することによって整理用バーコードを求める手段と、紙文書の表面のバーコードを読み取る手段と、指定されたバーコードを付与された文書を選択する手段とを有することを特徴とする文書情報システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は様式や書式の異なる複数種類の伝票（例えば、診療報酬請求明細書等）の読み取り、仕分け、保管、検索などを行なう伝票文書情報システムに関する。

【0002】

【従来の技術】文書の様式は、記載項目及び項目の順番を規定するが、その絶対的な位置座標、寸法を規定しない。文書の書式は、記載項目及び項目の順番を規定するだけでなく、その絶対的な位置座標、寸法を規定している。従来の光学的文字認識装置では、文書の書式をフォーマットデータとして予め設定しており、文書の各項目の絶対的な位置座標、寸法が規定されている紙文書を読み取ることができた。しかしながら、様式のみが規定されている文書、例えば、法令文書などを読み取ることはできなかった。

【0003】従来の伝票を読み取る文書情報システムは、特開昭56-11573号公報（特願昭54-87

338号 昭和54(1979)年7月10日出願、発明者：香川正幸、出願人：富士通株式会社)に述べられているように、伝票に予め入力すべき項目を指示するマーク記入領域を設け、当該領域に記入されたマークを読み取ることによつて、伝票上の所定領域を読み取る方式が提案されている。しかしながら、従来の方式では、マーク記入領域を設けているため当該記入領域の紙面上の占める面積が多くなり、入力項目を増やすことができないという問題があった。また、マーク記入領域を新たに用意する必要があるため、既存の各種伝票をそのまま読み取ることができないという問題がある。さらに、様式が法令等で決められている伝票文書では、従来のようにマーク記入領域を設けることは法令上様式が規定されているため、認められず、マーク記入領域のない当該様式文書を読み取ることはできないという問題があった。

【0004】また、従来の文書画像のファイル方式は、特開昭59-14070号公報(特願昭57-123460号 昭和57(1982)年7月15日出願、発明者：江尻公一、出願人：株式会社リコー)に述べられているように、中間調画像である写真や2値図形画像である棒グラフを含む文書の領域から特徴パラメータを抽出し、領域の画像と関連付けて記憶する画像ファイル方式が提案されている。しかしながら、従来の方式では、写真と棒グラフの識別を行なうことが述べられているだけであり、伝票のように罫線並びに罫線からなる枠と文字列からなる文書を識別することは述べられていない。また、伝票の文字列を文字認識することや、文書の領域を守秘の程度によってマスクするファイル方式について考慮されていない。

【0005】また、定形的な入力文書に対して文書の種類を自動的に識別し、すでに格納された文書と同一の分類コードを与えるファイリング方式が、特開昭61-75477号公報(特願昭59-196616号、昭和59(1984)年9月21日出願、発明者：中野康明、藤沢浩道、藤繩雅章、出願人：株式会社日立製作所)に述べられている。しかしながら、この従来方式では、画像を格納することが主であり、読み取り領域を設定し文字認識を行なうことや、当該認識結果のデータを基に記載内容の妥当性を検定することが考慮されていない。このため、文書の種類に応じて記載内容を読み取りかつ内容の妥当性を検定することができない。

【0006】

【発明が解決しようとする課題】上記各従来技術は、入力した紙文書そのものに対して整理番号若しくはバーコードを付与することが記載されておらず、蓄積した画像ファイルから所望の文書画像を表示、印刷することはできるが、原本である紙文書そのものを探索することは困難であり、原本を探す作業は多くの人手が必要な作業であった。また、紙文書そのものを仕分けすることが考慮されておらず、原本の分類、整理、収納に多くの人手を

要していた。

【0007】また、上記各従来技術では、蓄積した文書画像を表示、印刷する場合、守秘項目について考慮されておらず、例えば、文書にプライベートな内容や、ビジネス上重要な情報が記載されている場合でも検索結果の画像を画面にそのまま表示するか、若しくは、画像単位ごとに全面の表示を禁止するかのいずれかであり、文書の部分領域を守秘レベルに応じて見ることができなかつた。

【0008】また、上記各従来技術では、入力した紙文書はすべて蓄積することが前提であり、内容によって、取捨選択して蓄積することは考慮されていない。このため、大量の文書を蓄積することとなり、小規模のファイルシステムでは記憶装置の容量を超えるという問題があつた。

【0009】本発明の第1の目的は複数の伝票文書の様式の種類を識別し、読み取り領域を自動的に設定し文字認識を行い、記載内容の妥当性を検定することにある。

【0010】本発明の第2の目的は認識結果によって整理番号若しくはバーコードを紙文書に印刷し、紙文書の仕分けを行なうことにある。

【0011】本発明の第3の目的は文書画像から守秘内容の記載された領域を抽出し、守秘領域の画像をマスクして検索結果の表示、印刷を行なうことにある。

【0012】本発明の第4の目的は文書の記載内容によって蓄積の必要性の有無を自動的に判定し、記憶領域を節約することにある。

【0013】本発明の第5の目的はネットワークを介して認識、仕分け、蓄積を行なう文書情報システムを提供することにある。

【0014】

【課題を解決するための手段】上記第1の目的を達成するために、文書画像から様式の種類を識別して記載内容を読み取り、読み取り結果が記載項目ごとに登録している条件を満足するかどうかを判定するようにしたものである。

【0015】また、上記第2の目的を達成するためには、認識結果により仕分け機構並びに印刷機構を制御するようにしたものである。

【0016】また、上記第3の目的を達成するためには、文書画像から項目の記載された部分領域を自動的に抽出し、守秘領域に対してアクセスのためのパスワードを付与するようにしたものである。

【0017】上記第4の目的を達成するためには、記載内容例えば点数を読み取り、点数が一定値より大きい文書のみ、その表面画像を蓄積するようにしたものである。

【0018】上記第5の目的を達成するため、ネットワークに画像を圧縮して伝送するとともに、認識結果を伝送し、そのデータによってネットワークに接続された仕

分けステーション、蓄積ステーションを制御するようにしたものである。

【0019】

【作用】複数種類の伝票に対してその表面画像を入力し、伝票文書の様式の種類を識別する。そして、様式の種類に応じて読み取り領域を抽出し、文字認識を行なう。認識結果のデータを基に記載内容の妥当性を予め設定している判定条件に従い検定する。さらに、認識結果によって整理番号を紙文書に打番若しくはバーコードを印字し、文書の種類に従って紙文書の仕分けを行なう。また、文書の種類に応じて守秘内容の記載された領域を抽出し、守秘領域の画像をマスクして表示できるようにする。

【0020】

【実施例】以下、本発明を図面を用いて詳細に説明する。図1は本発明の一実施例である伝票情報システムの構成を説明する図である。

【0021】ネットワーク100には、入力・仕分けステーション103、蓄積ステーション101、認識ステーション105、修正ステーション106、検索ステーション107が接続されている。入力・仕分けステーション103は計算機より構成されている。当該ステーション103には入力・仕分け伝票文書の画像入力と伝票文書の仕分けを行なう入力・仕分け装置104が接続されている。当該装置には、文書を設定するホッパ、仕分け後の文書を格納するスタッカ及びその制御部があり、さらに、画像を入力するための光学的撮像素子、照明装置等の光学系が具備されている。ここでは、入力した画像をデータ圧縮し、当該ステーション103を介してネットワーク100に画像データを送出する。また、ネットワーク100から認識結果を受け取り、これを仕分け制御信号として文書の仕分けを行なう。蓄積ステーション101は計算機から構成され、光ディスクもしくはアレイディスク等からなる大容量記憶装置102が接続されており、認識結果のデータ並びに入力された画像の蓄積を行なう。蓄積ステーション101では、認識結果に基づいて蓄積すべき文書を選別することができる。また、守秘内容の記載されている部分領域に対して、アクセス情報を付与して当該記憶装置102に蓄積する。認識ステーション105では、ネットワーク100を介して送られてきた画像に対して、伝票文書の種類を求めるシート識別処理を行ない、項目に記載されている文字列の認識を行なう。修正ステーション106では、認識結果の修正を画面を用いて対話的に行なう。ここでは、読み取り領域を拡大して画面に表示し、修正を容易にしている。検索ステーション107では、蓄積ステーション101に蓄積された伝票画像をネットワークを介して検索することができ、項目に記載されている内容を手掛かりに所望の伝票を探索し、画面に表示することができる。この時、守秘内容の記載されている部分領域に対し

てアクセス情報が付与されており、使用者のアクセスのレベルにより当該守秘項目をマスクして表示することができる。

【0022】図2は伝票文書の処理フローを説明する図である。伝票文書の一例として、診療報酬請求明細書(レセプト)の処理フローを示す。レセプトに記載された主要項目についてデータを入力し、計算機により資格記録の確認処理を行なう。本処理フローでは、先ず、ステップ200においてレセプトを入力用のホッパーに設定する。次いで、ステップ201でレセプトの仕分け処理を行なう。ここでは、レセプトの種類として、外来分レセプト210と入院分レセプト211にその様式を識別して仕分けを行なう。ステップ202では、初期情報を入力する。ここでは、初期情報として、レセプトの処理を各県単位に行なうための地域属性として、自県と他県の区別を入力する。また、レセプトの処理を行なう事務所名、診療区分、整理番号の初期値を登録する。なお、入院分レセプト211については、整理番号をステップ208で打番する。次いで、ステップ203、ステップ204で読み取りを実行する。ステップ203は、伝票の一括処理を行なうバッチ読み取り処理であり、認識結果の正常分ならびに不読(リジェクト)分を別のスタッカーに格納するとともに、整理番号を印字する。ステップ204は、逐次修正読み取りの処理であり、認識の結果リジェクトされたレセプトに対して、修正画面を介して対話的に認識結果を修正する。ステップ203で認識された正常分のデータと、ステップ204で修正され正常となったデータはフロッピーディスク等に格納する。また、ステップ203でリジェクトとなったレセプトは、別のスタッカーに格納されることになる。なお、入院分レセプト並びにステップ203のリジェクト分をまとめてステップ209でパンチ作業を行なう。レセプトに記載の診療内容をもとに、ステップ207に示すように資格記録確認処理等の機械処理を行う。ここでは、高額な点数が記載されているレセプトのみを選択してステップ205で光ディスクに格納している。ステップ207の資格記録の確認処理では、ステップ205で光ディスクに格納されたレセプトの検索206を行ない、所望のレセプトを画面に表示して資格記録の確認を行なう。

【0023】図3は伝票文書の画像入力、蓄積、検索、内容点検を行なうシステムの処理フローである。ステップ300で伝票文書の表面画像を入力し、ステップ301で伝票の様式の種類を識別するシート識別処理を行なう。当該シート識別の詳細処理フローは図4にて後述する。次いで、ステップ302において、当該様式の種類に応じて読み取り領域を抽出し、読み取り領域を設定する。当該読み取り領域設定ステップ302の詳細処理フローは図9にて後述する。ステップ303では読み取り領域内において文字認識を行ない、ステップ304で当

該認識結果のデータを基に記載内容の妥当性を検定する。当該ステップ303の詳細処理フローは図10にて後述する。

【0024】図4は、ステップ301のシート識別処理の詳細処理フローを示す図である。ここでは、入力した画像から枠及び文字行を抽出し、予め標準的な文書から登録したシート識別用辞書430とのマッチングを行なう。先ず、ステップ400では入力画像から微小なノイズを除去する。次いで、ステップ401で黒画素の連続する線分であるランを用い野線を抽出し、ステップ402で当該野線の傾き角を算出する。ステップ403で画像を当該傾き角だけ回転させ、回転修正を行なう。そして再び、ステップ404で野線を抽出する。図6は、この一連の処理により画像から水平野線および垂直野線を取り出し、その始点座標、終点座標を算出する過程の説明図である。伝票画像600に対して、601は水平野線602、603、604を抽出した結果である。また、610は垂直野線611から616を抽出した結果である。ステップ405では文字のつながりである文字行を長方形の形で抽出し、その角の座標を求める。文書の水平並びに垂直方向の位置ずれを修正するため、ステップ406で文書の基準点を求める。ステップ407では、水平野線ならびに垂直野線から枠の角に相当する交点の座標を算出する。各野線は直線式で表現できるため、簡単に直線の交点座標を求めることができる。ステップ408では、求めた交点座標に対して、大きさの正規化を行なう。さらに、ステップ409では、交点を基に枠を生成する。図7は水平野線及び垂直野線から交点を求め、次いで枠を生成する過程を説明する図である。水平野線602～604、並びに垂直野線611～616をもとにその交点710～723を算出する。そして、交点を順次辿り枠701、702～706を生成する。図4のステップ410では、枠のマッチングを行なう。この時、シート識別用辞書430に格納されている複数個の標準の枠とのマッチングを行なう。当該辞書430には、枠位置及び文字行の位置が予め登録されており、その登録のフローは、先ず、ステップ420で標準の伝票画像を入力し、次いで、ステップ421で入力すべき枠及び文字行の位置を画面から指定する。ステップ422では位置並びに回転の補正を行ない、辞書として登録する。さらに、ステップ411では、文字行マッチングを行なう。この時、シート識別用辞書430に格納されている複数個の標準の文字行とのマッチングを行なう。文字行マッチングの詳細は図8において後述する。上に述べた枠マッチング410と文字行マッチング411によって求めた結果に従って、標準の伝票の内、もっとも類似している伝票が有するシート種類番号をステップ412で決定する。さらに、ステップ413で、生成した枠位置ならびに文字行の位置を出力する。また、ステップ414でシート種番号を出力する。

【0025】図5は位置ずれ・回転・伸縮を有する認識座標系を示している。伝票シート500に506で示す枠野線が存在している。認識座標系を伸縮の無い認識座標系(X_r, Y_r)511で、伸縮のある認識座標系(X_m, Y_m)512で表現する。これらの座標系は、センサと光学系からなる観測座標系である。また、フォーマットデータ生成用座標系(X_f, Y_f)513はシートの端を基準にした座標系であり、実物の寸法を表す座標系である。伸縮の無い認識座標系で表現したシート内の目標点(x_r, y_r)507は、認識座標系の原点510とフォーマットデータ生成用座標系の原点509とのずれ量(d_x, d_y)と回転角 θ により表現できる。

【0026】

【数1】

$$x_r = d_x + x_f \cdot \cos \theta + y_f \cdot \sin \theta$$

$$y_r = d_y - x_f \cdot \sin \theta + y_f \cdot \cos \theta$$
 ここで、(x_f, y_f)はフォーマットデータ生成用座標系で表した目標点507の座標である。さらに、伸縮のある認識座標系で表現したシート内の目標点507(x_m, y_m)は、伸縮率 m_x, m_y を、

【0027】

【数2】

$$x_m = m_x \cdot x_r$$

$y_m = m_y \cdot y_r$

とすると、

【0028】

【数3】

$x_m = m (d_x + x_f \cdot \cos \theta + y_f \cdot \sin \theta)$
 $y_m = m (d_y - x_f \cdot \sin \theta + y_f \cdot \cos \theta)$

と表すことができる。シート内の読み取りでは、上式に従って、フォーマットデータを変換して画像にアクセスし文字認識を行なう。また、位置合わせ用基準点508の認識座標系の値を伸縮の無い認識座標系(x_{r0}, y_{r0})、伸縮のある認識座標系(x_{m0}, y_{m0})で表す。また、フォーマットデータ生成用座標系の値を(x_{f0}, y_{f0})で表す。位置ずれ量(d_x, d_y)は、

【0029】

【数4】

$$d_x = (1/m_x) \cdot x_{m0} - x_{f0} \cdot \cos \theta - y_{f0} \cdot \sin \theta$$

$$d_y = (1/m_y) \cdot y_{m0} + x_{f0} \cdot \sin \theta - y_{f0} \cdot \cos \theta$$

で表すことができる。回転角 θ は野線の傾きを基に検出する。次いで、伸縮率 m_x, m_y を求めるが、これは、例えば、一番外の黒線枠の横幅、縦幅を検出し、フォーマットデータの該当する横幅、縦幅と比較して算出する。さらに、位置合わせ用基準点の認識座標系における位置(x_{m0}, y_{m0})を求め、フォーマットデータに予め設定している当該点の位置(x_{f0}, y_{f0})（フォーマットデータ生成用座標系で表現）と比較して、位置ずれ量(d_x, d_y)を求める。シート識別では、伸縮のある認識座標系において検出した目標点の座標(x

$m, y m$) をフォーマットデータ生成用座標系 ($x f, y f$) に変換し、これをシート識別用辞書の書式の座標データと比較する。

【0030】文字行の相対的な位置関係を表現する方式を図29に示す。今、注目する文字行2900に対して、8方向の位置関係を表現している。方向を東西南北で表すと、注目する文字行2900の北方向2901には文字行2911がある。また、南方向2905には文字行2915が、東方向2903には文字行2913が、西方向2907には文字行2917が、それぞれある。さらに、注目する文字行2900の北東方向2902には文字行2912が、北西方向2908には文字行2918が、南東方向2904には文字行2914が、南西方向2906には文字行2916が、それぞれある。このように、注目する文字行に対して、8方向の位置関係を有する文字行を表現する。

【0031】図8はステップ411の文字行マッチングの処理方式を説明する図である。伝票文書の文字行の部分領域の例を810、811、812に示す。文字行領域810では、抽出した文字行を801、802、803示しているが、これらの相対的な位置関係は上下の南北関係になっている。いま、注目する文字行を801とすると、文字行802は文字行801に対して南方向820にある。また、文字行802に対して、文字行803は、同じく、南方向821にある。一方、文字行領域811の例では、文字行804、805、806の相対的な位置関係は左右の東西関係である。文字行804に対して、文字行805は東方向822にあり、文字行805に対して806も又、東方向823にある。さらに、文字行領域812の例では、斜め方向の位置関係にあり、文字行807に対して、文字行808は北東方向824にある。また、文字行808に対して文字行809は、同じく、北東方向825にある。ステップ411の文字行マッチングでは、このような文字行の相対的な位置関係が予めシート識別用辞書430には登録されており、入力画像から抽出した文字行の相対的な位置関係と合致するかどうかを判定する。

【0032】図9は図3のステップ302で示した読み取り領域設定処理の詳細処理フローである。先のシート識別のステップ301で得られたシート種番号をステップ900で入力する。次いで、ステップ901で読み取り領域ごとの処理に移り、ステップ902でシート種類ごとに登録されているフォーマットパラメータ910から、該当する読み取りのフォーマットパラメータを読み込む。ここで、フォーマットパラメートとしては、文字種、文字数、読み取りフィールドの位置が具備されている。ステップ903では、フォーマットパラメータから読み込んだ読み取りフィールド位置とステップ406で求めた基準点位置からの位置ずれ量を基に、読み取り領域を算出する。

【0033】図10はステップ303で示した読み取り領域内文字認識処理の詳細処理フローである。ステップ1000で読み取りフィールドごとに、ステップ1001から1004までの処理を行なう。ステップ1001では読み取り領域内の文字行を抽出する。ステップ1001では、予め、読み取り領域内に印刷されている不要文字を除去する。そして、ステップ1003で抽出した文字行を一文字単位に分離する。ステップ1004では、一文字ごとの文字認識を行なう。この文字認識は、従来技術を用いて容易に実現できる。

【0034】図11は伝票文書を読み取り、当該読み取り結果をもとに伝票文書を選択して蓄積する伝票文書情報システムの処理フローを示す図である。ステップ1100で文書がつくるまで、ステップ1101以下の処理を繰り返す。ステップ1101で伝票文書の画像入力を行なう。次いで、ステップ1102で、ステップ301と同様のシート識別処理を行なう。次いで、ステップ1103で同じく、読み取り領域を抽出し、ステップ1104で読み取り領域内において文字認識を行なう。ステップ1105で、認識結果の数値データが予め設定された範囲内にあるかどうかを判定し、もし設定された範囲内にある場合は、ステップ1106で光ディスク等の記憶装置に当該画像を蓄積する。もし、設定された範囲内にない場合は、記憶装置に当該画像を蓄積しない。上述の一連の処理をステップ1100で文書が尽きるまで行なう。この処理により入力した紙文書はすべて蓄積する必要はなく、記載内容によって、取捨選択して蓄積することができる。このため、小規模のファイルシステムでは記憶の容量を節約できるという効果がある。

【0035】伝票文書の画像入力、蓄積、検索、内容点検を行なう文書情報システムにおいては、使用者のアクセスレベルに応じて守秘領域の画像をマスクして表示する。図12は伝票文書画像から守秘領域を登録する処理フローを示す。ステップ1200で画像を入力し、ステップ1201でシート識別処理301を行なう。そして、ステップ1202で守秘領域を抽出する。ここでは、先のシート識別のステップ1201でシート種と抽出した領域の位置を出力しており、領域ごとに守秘領域であるかどうかを判定する。この時、守秘領域であるかどうかは、予め登録しているシート識別用辞書430を読み取るものとする。当該辞書430には、枠で示された領域ごとに守秘のレベルが具備されている。ステップ1203で守秘領域の座標を画像データとともに記憶装置に格納する。

【0036】図13は伝票画像に対して、使用者のアクセスレベルにより守秘領域の画像をマスクして表示する処理のフローである。先ず、ステップ1300で使用者のパスワードを入力する。そして、ステップ1301で使用者のアクセスレベルを図16に示す使用者アクセスレベルテーブル1600を用いて設定する。図16は使

用者アクセステーブルの説明図である。使用者名1601、パスワード1602に対して、アクセスレベル1603が格納されており、当該使用者に対応するアクセスレベルを設定する。次いで、ステップ1302で領域がつくるまで、ステップ1303、1304の処理を繰り返す。ステップ1303では、領域ごとに設定されているアクセスレベルを領域アクセステーブル1500から読み出す。図15は領域アクセステーブルを示す。当該テーブル1500は、標準の伝票文書に対して、領域に付与された識別番号1501ごとにアクセスレベル1502が付与されている。そして、ステップ1304で使用者のアクセスレベルが領域のアクセスレベルより強いかどうかの判定を行なう。もし、使用者のアクセスレベルが強い場合は、ステップ1305で当該領域のマスクを解除する。一方、その反対に、領域のアクセスレベルの方が使用者のアクセスレベルより強い場合は、ステップ1306で当該領域にマスクを設定する。以上の一連の処理をステップ1302で領域がつくるまで繰り返し行なう。そして、ステップ1307で伝票画像を表示する。ここでは、領域ごとに表示処理を実行し、領域にマスクが設定されている部分画像はそのまま表示せず、例えば、黒画像により塗りつぶした画像を表示する。領域のマスクが解除されている部分画像はそのまま表示する。本実施例では、蓄積した文書画像を表示、印刷する場合、守秘項目についてアクセスレベルが付与されており、例えば、文書にプライベートな内容や、ビジネス上重要な情報が記載されている場合に、文書の部分領域を守秘レベルに応じて見ることができ、機密を保つことができるという効果がある。

【0037】図14は伝票文書画像の表示の一例である。表示画面1400において、伝票の各項目に対応した領域1401、1402、1403、1404、1415、1405、1406、1407、1408、1409が設定される。この時、使用者のアクセスレベルに応じて、斜線部分1410、1411、1412、1413、1414において、守秘のためにマスクが設定されており、当該領域は黒く塗りつぶされることになる。

【0038】図17は伝票文書の読み取り結果によってバーコードを当該文書に印刷する処理フローを示す。ステップ1700で文書画像を入力する。入力した画像に対して、ステップ1701で伝票文書の種類を識別するシート識別を行ない、ステップ1702で文字認識によりシート内の領域の読み取りを行なう。そして、ステップ1703で当該認識結果に従ってバーコードの印字を紙文書の表面に行なう。

【0039】図18はバーコードにより文書を探索する処理フローを示す。ステップ1800で、探索したい文書の整理番号を入力すると、ステップ1801で整理番号に該当するバーコードを設定する。そして、ステップ1802で文書が尽きるまで、ステップ1803、18

04、1805の処理を繰り返す。先ず、ステップ1803では、伝票文書の紙送りを実行し、紙表面に印刷されているバーコードをステップ1804で読み取る。次いで、ステップ1805において、読み取ったバーコードがステップ1801で設定されたバーコードと一致するかどうかを判定する。もし、一致すれば、所望の伝票文書であるとして、ステップ1806で選択スタッカに格納する。もし、ステップ1805で一致しなければ、ステップ1807で通常のスタッカに格納し、ステップ1802で文書がつくるまで処理を繰り返す。本実施例により原本にバーコードを印刷しているため、蓄積した画像ファイルから所望の文書画像を表示、印刷するだけでなく、原本である紙文書そのものをバーコードを用いて容易に探索することができるという効果がある。図19は伝票文書の種類を識別し仕分けを行なう処理フローである。ステップ1900で伝票文書画像を入力し、シート識別用辞書1910によってステップ1901でシート識別を行なう。次いで、ステップ1902でシート内読み取りを実行する。この時、予め登録しているフォーマットパラメータ1911と文字認識用辞書1912を用いる。ステップ1903で、シート識別の結果であるシートの種類によって伝票文書の仕分けを行なう。本実施例により、従来人手により行なわれていた伝票文書の仕分けを自動的に行なえるようになり、人手による作業量を低減することができる。

【0040】図20はネットワークに接続された入力・仕分けステーションと認識ステーションの動作制御を説明する図である。ネットワーク2000には入力・仕分けステーション2002と認識ステーション2001が接続されている。処理の制御は、先ず、認識ステーション2001から全体システムの起動を入力・仕分けステーション2002に送る。入力・仕分けステーション2002、この起動指令を受け、ホッパに設定された伝票文書は入力装置2004のスキャナで画像を入力する。画像は入力装置2004内にある画像プロセッサにおいてデータ圧縮され、ネットワーク2000を介して認識ステーション2001に転送される。転送された画像は伸張され、シート識別と読み取りを行なう。認識結果は、ネットワーク2000を介して入力・仕分けステーション2002に転送され、仕分け装置2005のスタッカを選択する。上記の処理をホッパの設定されている伝票文書がなくなり、ホッパが空になるまで行なう。なお、スタッカが途中で満杯になった場合は、警告音を発生し、一時停止する。ホッパが空になれば、入力・仕分けステーション2002から終了指令を認識ステーション2001に送信し、全体システムの処理が終了する。

【0041】図21は伝票文書の種類を識別し文字認識を行い認識結果によって整理番号を紙文書に打番して仕分けを行なう伝票文書処理のフローである。ステップ2100で伝票文書の画像を入力する。ステップ2101

で伝票の種類を識別するシート識別を実行する。ステップ2102でシート内読み取りを行ない、当該認識結果によってステップ2103で整理番号を打番する。ステップ2104では伝票の種類、若しくは、記載内容によって仕分けを行ない所定のスタッカに格納する。

【0042】図22は整理番号を打番した伝票文書の一例である。伝票文書2200において、2201で示す整理番号を打番している。文書の端部の位置を検出しているため、当該整理番号は予め設定している座標位置に打番することができる。また、打番のためのインク色を指定色に切り換えることができる。

【0043】図23は入力・仕分けの構成を示す図である。ホッパ2300に伝票文書2320が設定されている。2304は計算機であり入力・仕分け全体を制御するとともに。ネットワーク2321との通信制御を行う。ホッパ制御部2310は、ホッパ機構部2315を制御し、順次、紙をセンサ部2313に送る。センサ制御部2311は画像信号をデジタル化し計算機2304に送る。2314は打番機構であり、打番制御部2312によって計算機2304から送られてきた整理番号を印字する。仕分け機構部2316、2317、2318は計算機2304の指令により、送られてきた伝票文書を所定のスタッカに格納するよう制御されている。スタッカ2301、2302、2303には選択された伝票文書が順次格納されている。各スタッカには文書の満杯を検知するセンサ2307、2308、2309が具備されており、満杯検知センサ制御部2306によってスタッカ満杯信号が計算機2304に送られる。2305は警告音発生部であり、スタッカの満杯が検知されると、警告音を発生し、計算機2304は処理を一時停止する。

【0044】図24は伝票文書情報システムの画面の遷移図の一例である。システム立ち上げ時のメイン画面2401を図25に示す。メイン画面2500は、選択項目2501並びに選択項目の入力プロンプト2502を具備している。入力方法指定画面2402によって、2種類の入力を切り換えることができ、その一つはスキャナ入力画面2403、もう一つはファイル入力画面2404で行なう。2405は画像表示画面であり、指定した入力方法で画像を入力し、その入力画像を当該画面で表示確認することができる。当該画面から、メイン画面2401に戻ることができ、例えば、様式／書式識別画面2406に遷移する。ここでは、図26に示すように、抽出した枠2601を画面2600に表示することができる。また、画面2602に示すように、抽出した枠2603と辞書として具備していた枠2604を重ねあわせて表示することもできる。2408は認識結果画面であり、文書画像と認識結果を同時に表示し、修正を行なうことができる。図27は認識結果画面を説明する図である。伝票文書2700に対して、上部画像表示エ

リア2701と下部画像表示エリア2702を拡大して表示する。認識結果画面2710において、2711に上部画像を拡大して表示している。また、2712に下部画像を拡大して表示している。認識結果は2713の部分に表示しており、修正を対話的に行なうことができる。図28は、認識結果画面の他の実施例を説明する図である。本実施例では、認識結果画面を2画面として、第1画面2800には上部画像2801を表示し、認識結果は2802の部分に表示している。また、第2画面2803には、下部画像2804を表示しており、同じく認識結果を2805に表示し、対話的に修正を行なう。

【0045】図30はステップ304の妥当性検定の詳細処理フローであり、レセプトを対象とした例である。ステップ3000で、性別と請求内容の検定を行なう。性別によって医療行為が異なるものがあり、性別により検査や手術等の請求内容が該当するかどうかを検定する。さらに、ステップ3001では、年齢と請求内容の検定を行なう。ここでは、年齢により調剤や手術の内容に制限があり、請求内容が当該年齢において適切であるかの検定を行なう。さらに、ステップ3002では、年齢と伝票様式の検定を行なう。ここでは、年齢により、例えば、新生児、乳幼児、老人等により伝票の様式が予め規定されており、シート識別で求めた伝票の様式が記載されている年齢と合致するかどうかの検定を行なう。また、ステップ3003で保険証番号と登録されている氏名との一致を判定する。さらに、ステップ3004と3005において、診療実日数及び合計点の検定を行なう。

【0046】

【発明の効果】文書画像から様式の種類を識別して記載内容を読み取り、読み取り結果が記載項目ごとに登録している条件を満足するかどうかを判定するようにしておらず、複数の伝票文書を読み取ることができる。また、読み取り領域を自動的に設定し文字認識を行い、記載内容の妥当性を検定しており、伝票の記載内容を自動的に点検することができる。認識結果により仕分け機構並びに印刷機構を制御するようにしておらず、認識結果によって整理番号若しくはバーコードを紙文書に印刷し、紙文書の仕分けを自動的に行なうことができ、仕分け作業を低減するという効果がある。守秘領域に対してアクセスのためのパスワードを付与するようにしておらず、守秘領域の画像をマスクして検索結果の表示、印刷を行なうことができ、システムの信頼性を向上させる効果がある。記載内容例えば点数を読み取り、点数が一定値より大きい文書のみ、その表面画像を蓄積するようにしておらず、記憶領域を節約できるという効果がある。

【画面の簡単な説明】

【図1】伝票情報システムの構成を説明する図である。

【図2】伝票文書の処理フローを説明する図である。

【図3】伝票文書の画像入力、蓄積、検索、内容点検を行なうシステムの処理フローを示す図である。

【図4】シート識別処理の詳細処理フローを示す図である。

【図5】位置ずれ・回転・伸縮を有する認識座標系を説明する図である。

【図6】罫線座標を算出する過程の説明図である。

【図7】罫線交点算出並びに枠生成の処理過程を説明する図である。

【図8】文字行マッチングの処理方式を説明する図である。

【図9】読み取り領域設定処理の詳細処理フローの説明図である。

【図10】読み取り領域内文字認識処理の詳細処理フローの説明図である。

【図11】伝票文書の選択蓄積を行なう伝票文書情報システムの説明図である。

【図12】伝票文書画像から守秘領域を登録する処理フローの説明図である。

【図13】使用者のアクセスレベルにより守秘領域の画像をマスクして表示する処理フローの説明図である。

【図14】伝票文書画像の表示の一例を説明する図である。

【図15】領域アクセステーブルの説明図である。

【図16】使用者アクセスレベルテーブルの説明図である。

【図17】認識結果を用いバーコードを文書に印刷する処理フローの説明図である。

【図18】バーコードにより文書を探査する処理フローの説明図である。

【図19】伝票文書の種類を識別し仕分けを行なう処理フローの説明図である。

【図20】ネットワークに接続されたステーションの動作制御の説明図である。

【図21】認識結果によって整理番号を打番し仕分けを行なう伝票文書処理フローの説明図である。

【図22】整理番号を打番した伝票文書の一例を示す図である。

【図23】入力・仕分けの構成を示す図である。

【図24】伝票文書情報システムの画面の遷移図である。

【図25】メイン画面を示す図である。

【図26】枠の表示画面を示す図である。

【図27】認識結果画面を説明する図である。

【図28】認識結果画面の他の実施例を説明する図である。

【図29】文字行の相対的な位置関係を説明する図である。

【図30】妥当性検定の詳細処理フローである。

【符号の説明】

100…ネットワーク、101…蓄積ステーション、105…認識ステーション、103…入力・仕分けステーション、301…シート識別処理ステップ、304…妥当性検定ステップ、410…枠マッチングステップ、411…文字行マッチングステップ、910…フォーマットパラメータ、1306…領域のマスク設定、1703…バーコード印刷ステップ、1804…バーコード読み取り、2103…整理番号打番、2104…仕分け処理。

【図7】

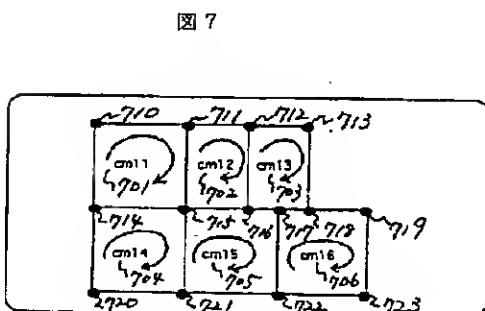


図7

【図10】

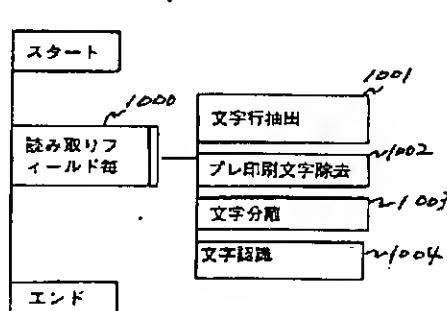


図10

【図17】

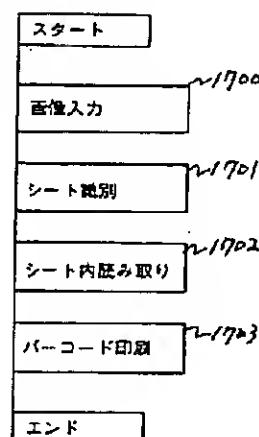
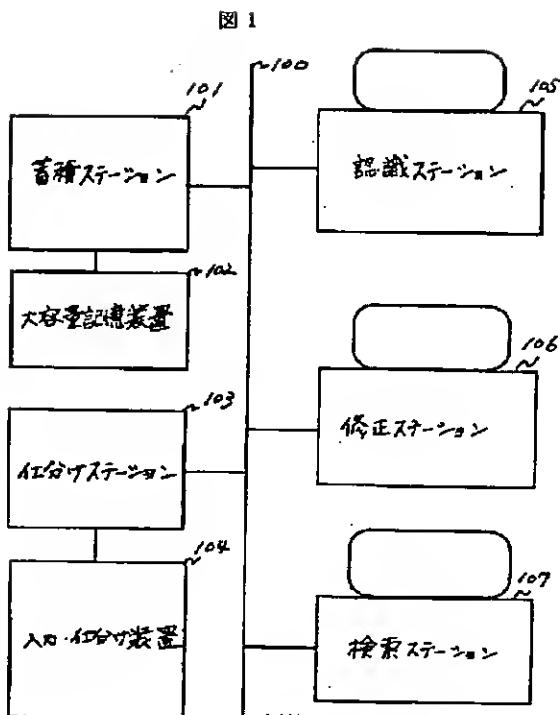


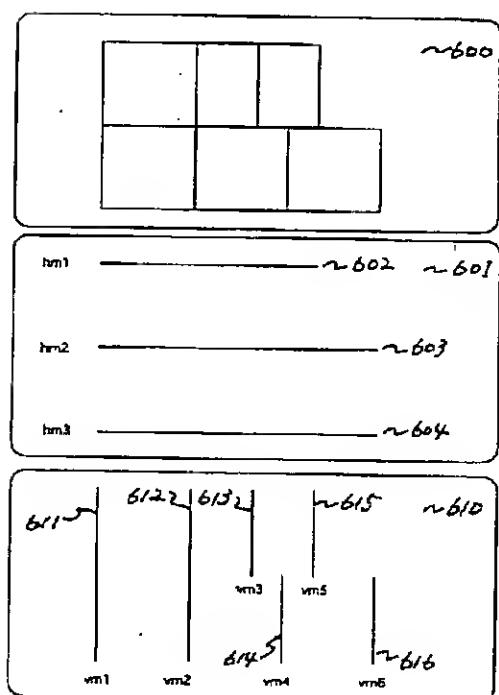
図17

[图 1]

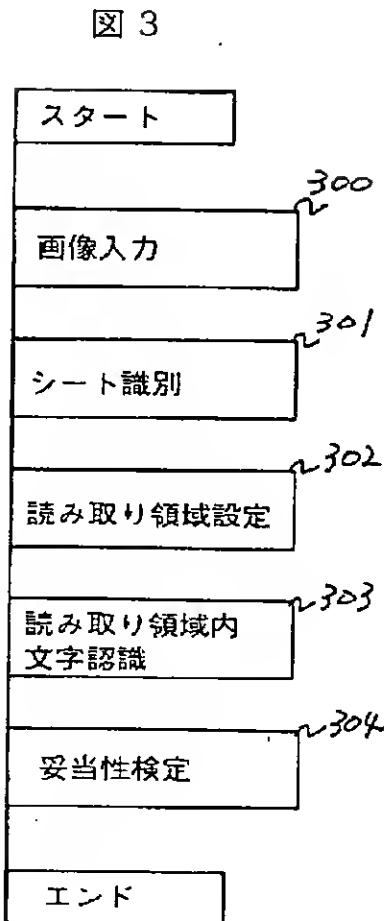


[図6]

5



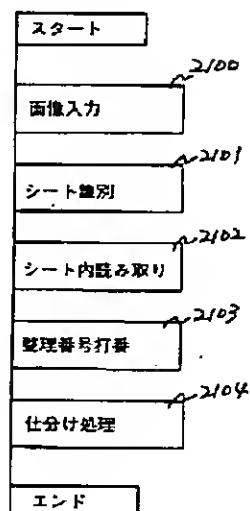
{图31}



[图 14]

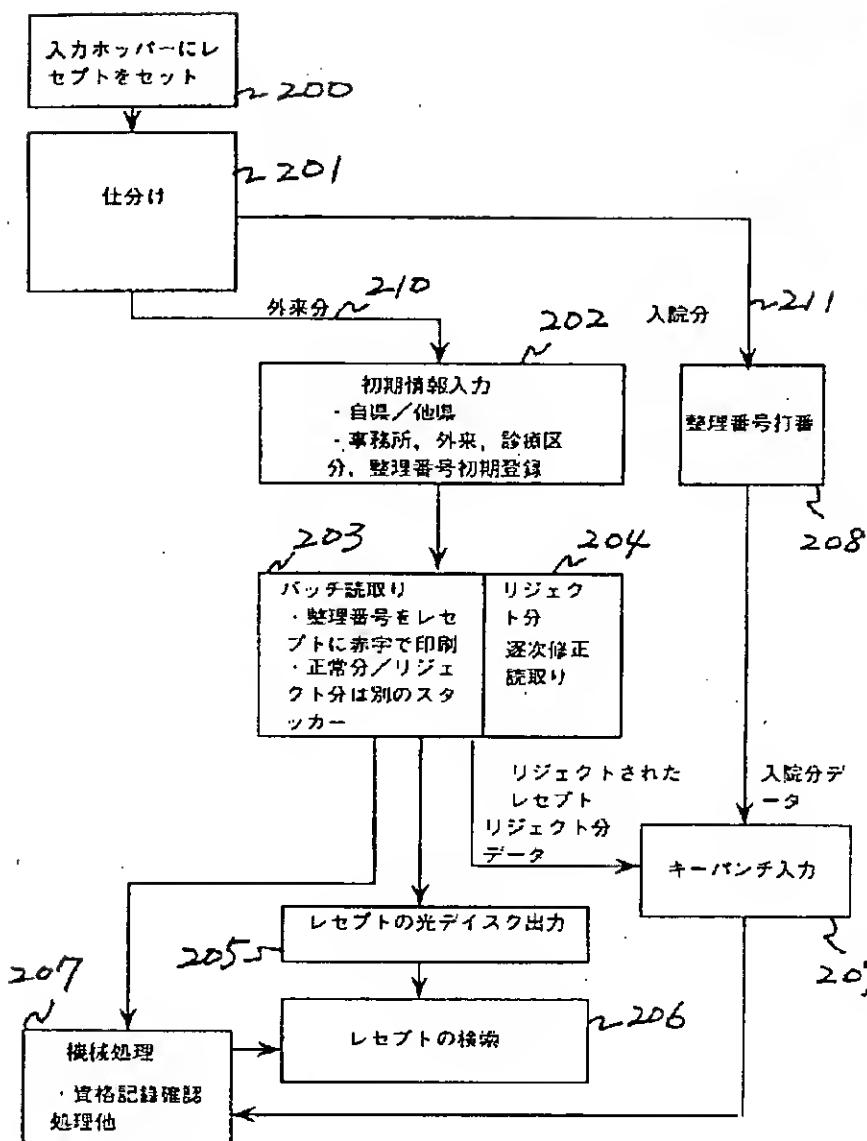
图 1-4

[図21]



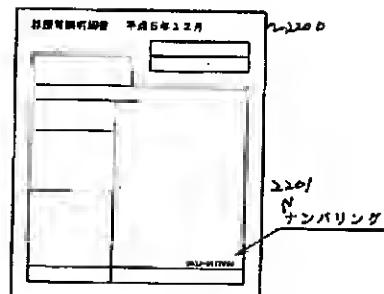
【図2】

図2



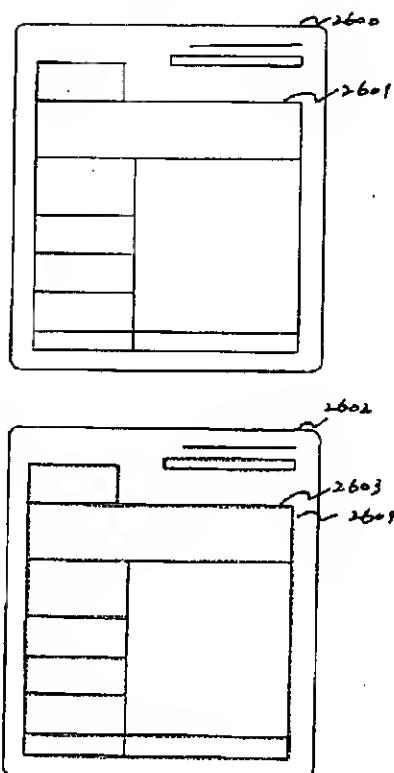
【図22】

図22



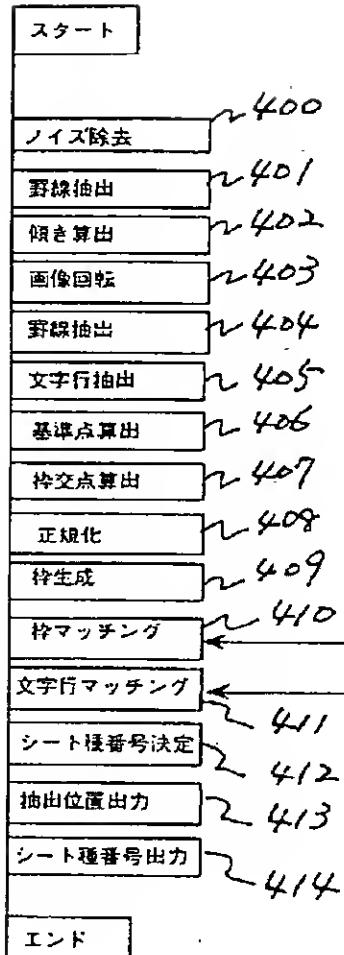
【図26】

図26



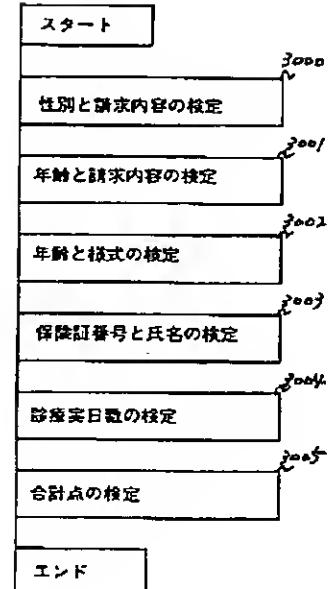
【図4】

図4



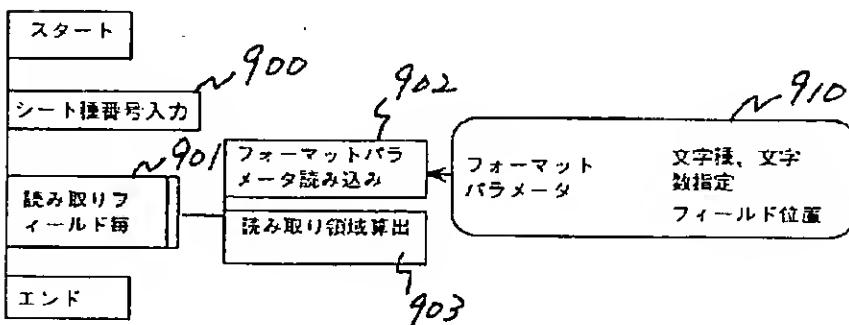
【図30】

図30



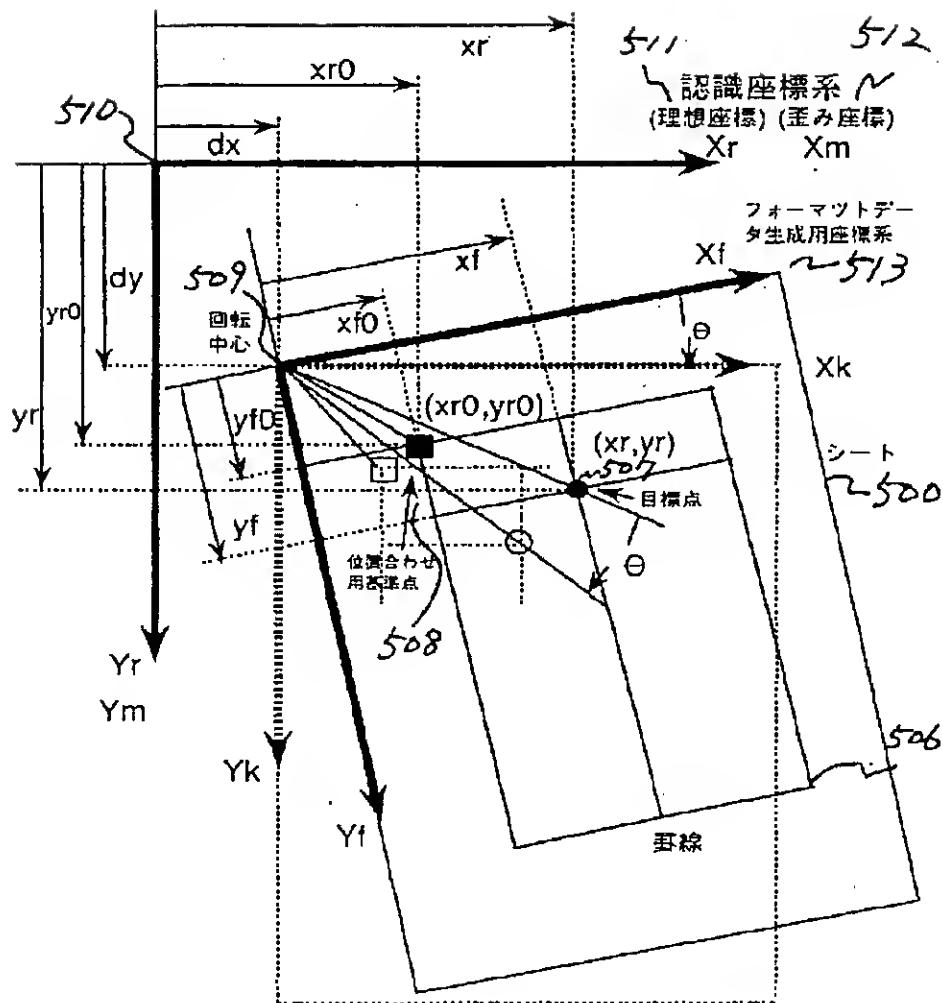
【図9】

図9



[図5]

5



[☒ 1 5]

[図16]

215

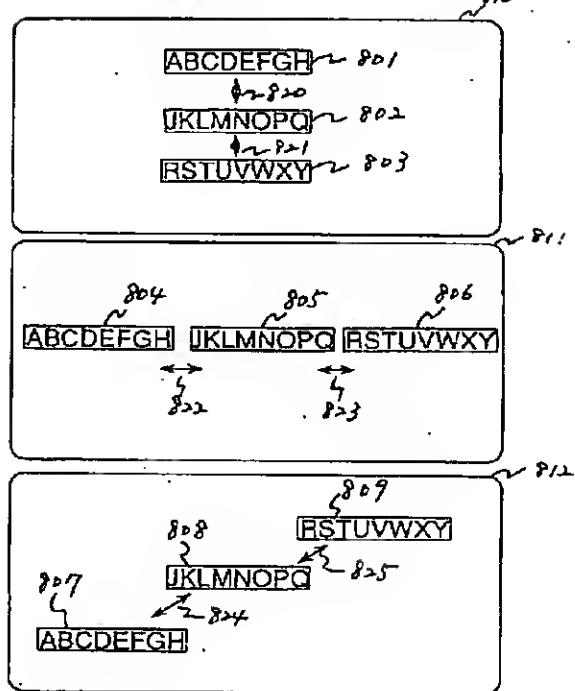
图 16

領域番号		1501	アクセスレベル	1502	1503
1		0			
2		1			
3		3			
4		2			

使用者名	パスワード	アクセスレベル
shima	abc	0
marukawa	bcd	2
koga	cde	3
nakashima	def	4

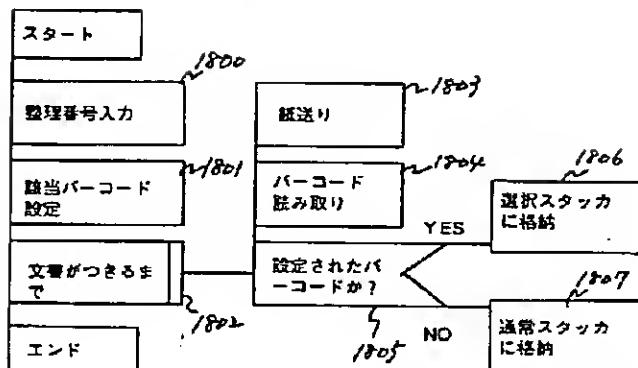
【図8】

図8



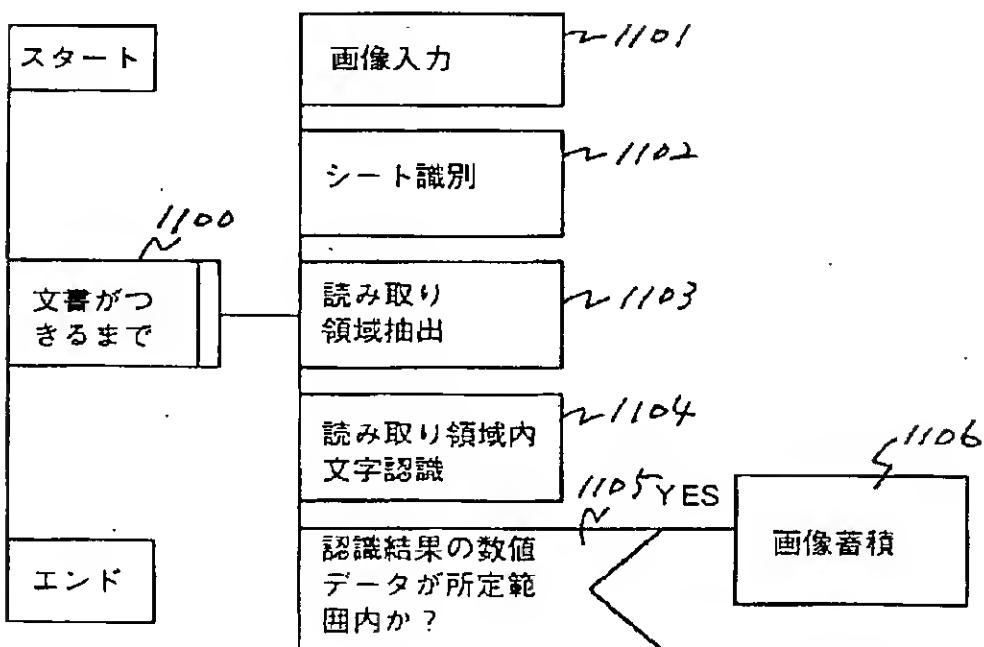
【図18】

図18



【図11】

図11



[図12]

[図19]

図 1 2

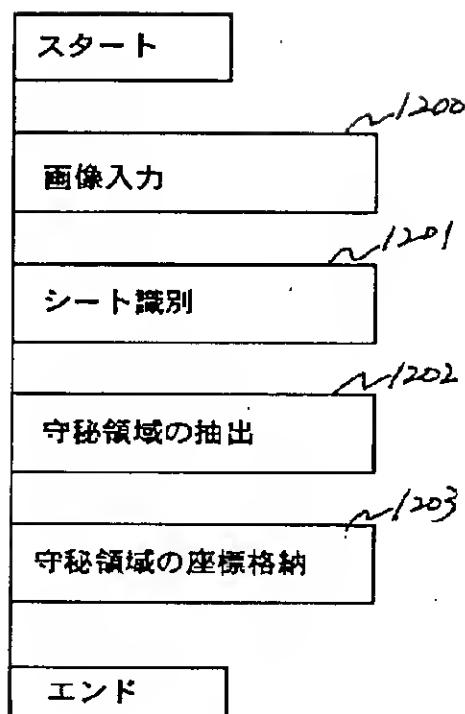
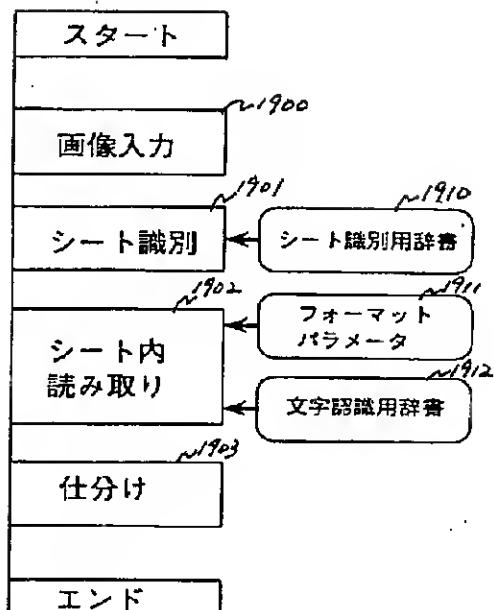


図 18

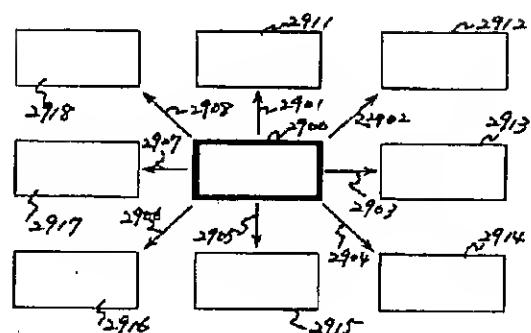
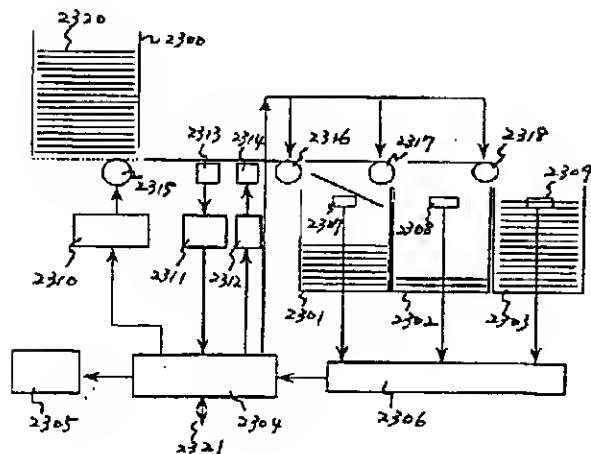


[图 29]

图 2-9

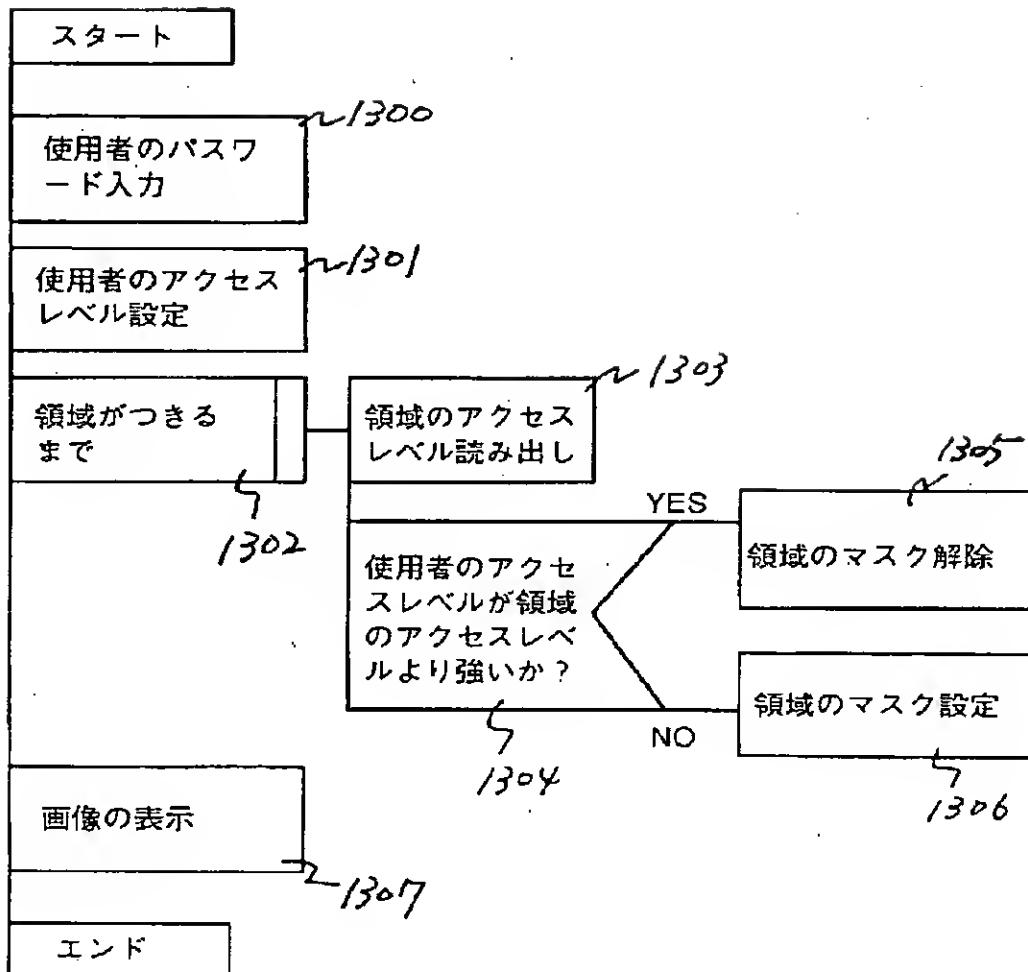
[图 2-3-1]

図 23



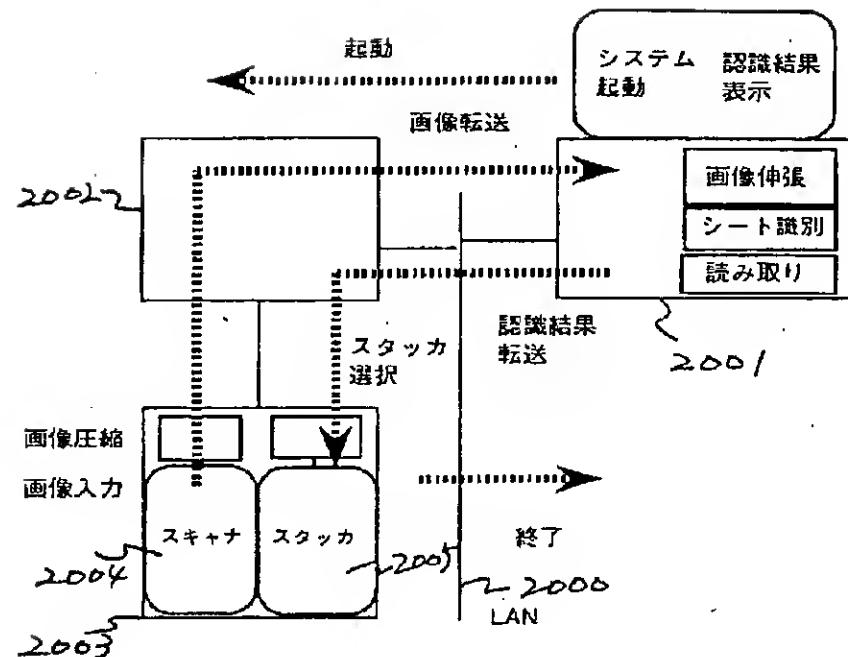
[図13]

図13



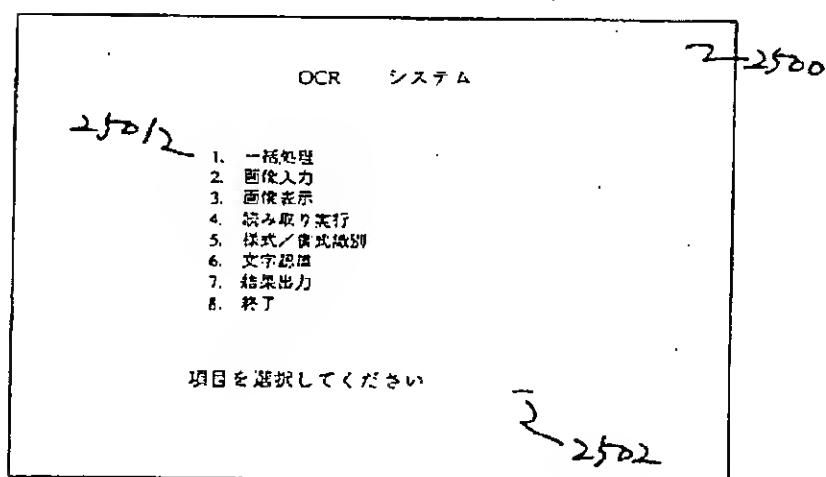
【図20】

図20



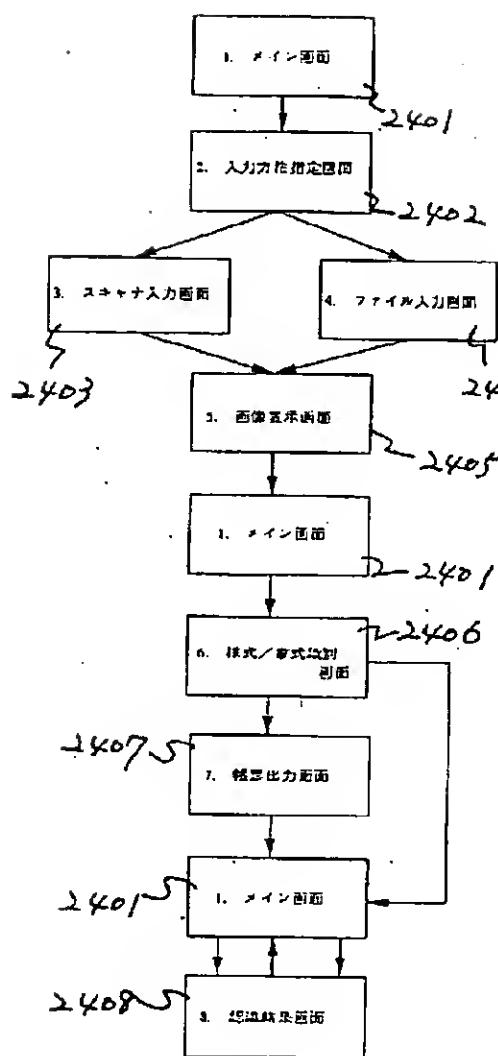
【図25】

図25



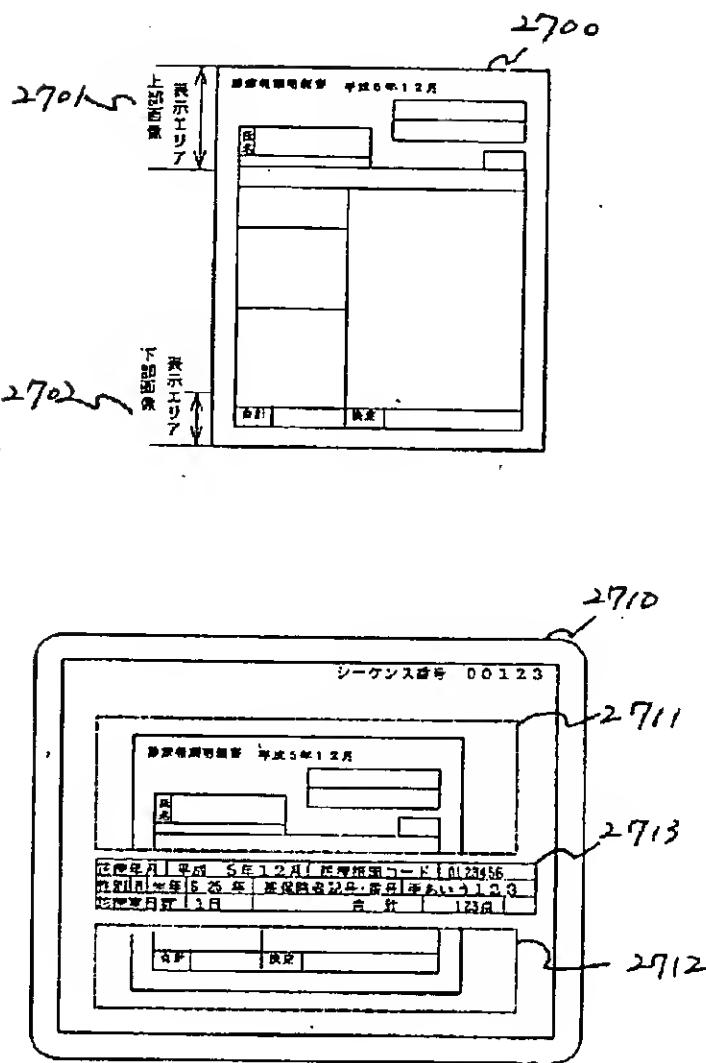
[図24]

図24



[図27]

図27



[図28]

図28

2801 上部画像

シーケンス番号 00123

記録期間開始 平成6年12月

氏名

記録年月 平成 6年12月	記録時間コード 0123456
性別 男年 52歳	記録件数記号 並び 00123
記録次日月 1日	合計 128点

2802

2804 下部画像

シーケンス番号 00123

会員 決定

記録年月 平成 6年12月	記録時間コード 0123456
性別 男年 52歳	記録件数記号 並び 00123
記録次日月 1日	合計 128点

2805

フロントページの続き

(72)発明者 古賀 昌史

東京都国分寺市東恋ヶ窪1丁目280番地
株式会社日立製作所中央研究所内

(72)発明者 中島 和樹

東京都国分寺市東恋ヶ窪1丁目280番地
株式会社日立製作所中央研究所内

(72)発明者 門田 彰三

神奈川県小田原市国府2880番地 株式会社
日立製作所ストレージシステム事業部内

(72)発明者 栗野 清道

神奈川県小田原市国府2880番地 株式会社
日立製作所ストレージシステム事業部内

(72)発明者 杉本 建行

神奈川県小田原市国府2880番地 株式会社
日立製作所ストレージシステム事業部内

THIS PAGE BLANK (USPTO)